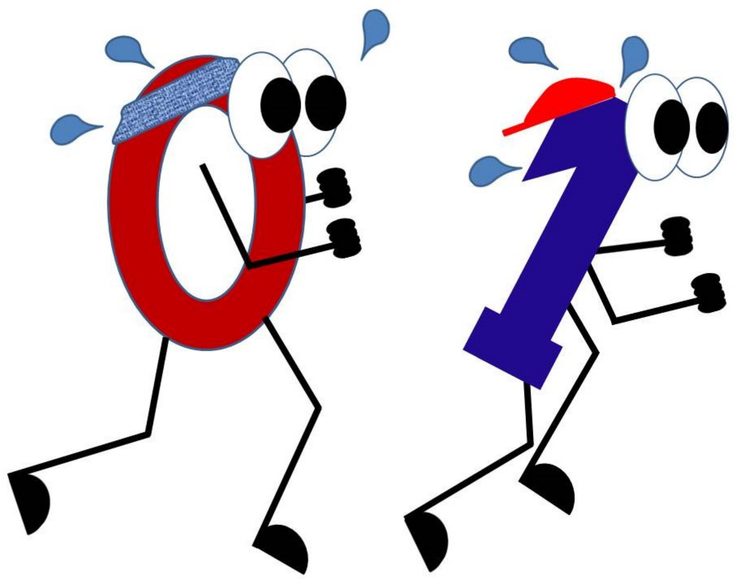
Science & Technology Facilities Council

www.stfc.ac.uk

Scientific Computing Training for NERC Researchers

Alison Pamment[1], Mat Evans[2], James Groves[3], Stephen Pascoe[1], Sam Pepler[1], Ag Stephens[1], Dan Walker[3], Louise Whitehouse[3]
[1]NCAS/BADC, Science and Technology Facilities Council, [2]NCAS/University of York, [3]NCAS/University of Leeds



Are your data management skills fit for the future?

Get into training with the NCAS Introduction to Scientific Computing course …

A team of NCAS staff, led by Professor Mat Evans, gave a pilot presentation of the "Introduction to Scientific Computing" course at the University of York, March 17th – 20th 2014. The course work included many practical exercises, carried out using laptop computers loaned by NCAS. Course administration was provided by NCAS staff from the University of Leeds. The course is designed to provide scientists with the building blocks needed to manage and analyse data effectively.

The main course modules are:-

- The linux shell
- Python programming basics
- Data analysis and visualisation tools in Python
- Automated logging of instrumental data
- Standard data file formats and metadata conventions
- CEDA archive services
- Introduction to the JASMIN computing facilities
- Software version control
- Introduction to parallelising your code

Our approach is to reuse, wherever possible, training materials that already exist in the public domain. Materials developed by the Software Carpentry organisation http://software-carpentry.org/) are especially useful in this context. Additional material has been developed by NCAS staff from CEDA (Centre for Environmental Data Archival, STFC) and the universities of York and Leeds. The full course materials can be viewed at http://www.ceda.ac.uk/ncas-york-2014/.

Why good data management skills matter

• Data are often a primary product of scientific research. We all have a responsibility to ensure that data produced from publicly funded research are preserved and made available for others to use

• Whether dealing with large quantities of numerical model output (big data) or many small files containing observations, having the ability to automate routine data processing tasks can save significant amounts of time

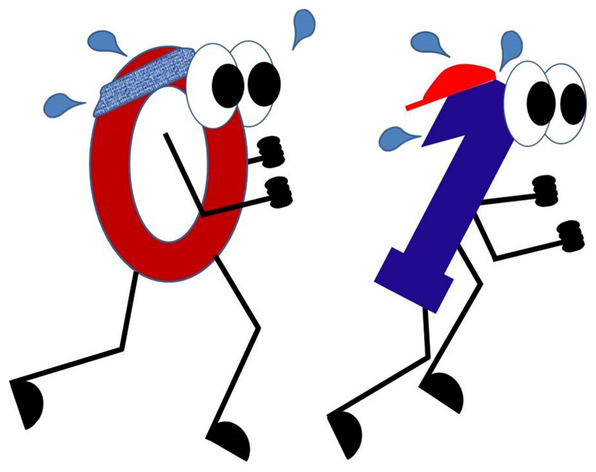
• Using standard file formats and accepted metadata conventions helps to ensure that your data will remain accessible and readable for many years to come

• A Digital Object Identifier (DOI) can be issued for datasets that satisfy certain conditions, making them easy to cite and increasing impact

• It is NERC policy that scientists produce an Outline Data Management Plan when applying for funding and a full Data Management Plan if the application is successful

The next course will be held at the University of Reading in the autumn of 2014.

For further details please contact: Alison.Pamment@stfc.ac.uk



National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

software carpentry